INBORN ERRORS OF METABOLISM

There is a class of disease known as “inborn errors of metabolism” that occur when an individual lacks a specific enzyme within one of the central metabolic pathways. Babies are often screened for this type disease at birth, and the fortunate ones can be helped (sometimes) by restrictive diets. Aside from the quality of life issues faced by those that suffer from these diseases, some metabolic diseases have historical significance (i.e. vampirism) or are a physical curiosity (i.e. coloration of Siamese cats). Because one of the fundamental concepts we have studied in this class is the process in which genetic information is converted into a protein sequence, it is worth exploring the basis of these genetic diseases.

The basis of this assignment is to discover the functional significance of a DNA coding sequence and its connection to an inherited metabolic disease. You will be assigned a genetic sequence, use bioinformatics tools to search for the function of the gene product, and finally report your findings in a short paper no longer than 3 pages in length. The paper is to be a specific type of document referred to as a Review paper. Scientific review papers present a comprehensive overview about a specific topic, either from a historical viewpoint or from current research. They summarize the information found in the scientific literature in order to communicate what scientists currently believe to be true about a topic.

Review papers are more flexible in their structure than a lab report, but still have some recognizable organization. For this assignment, there are two organizational structures you might consider. One way is to introduce the gene sequence you have been assigned and then gradually broaden the scope to describe the disease. The other option is to start more broadly describing the disease and then focus in on the actual gene sequence. In either case, you will need to include at least 3 references for the paper, one of which may be the internet site used for the gene sequence analysis. The other two references must be non-web page sources; good examples include Scientific American, the journal Science, or Proceedings of the National Academy of Science. Remember that Mary Iber is a valuable resource for help in finding reference materials.

Rules for the paper

1. The written portion of the paper may not exceed 3 pages, 12 pt. font, double spaced (about 1200 words).
2. Use at least 3 references; 2 must be scientific journals, and the sequence database must also be cited.
3. The bibliography or any diagrams are not included in the page count.
4. Information about the disease that is to be discussed:
   - the normal function of the affected enzyme—substrates, products, and the pathway to which it belongs.
   - the disease symptoms and why they occur (e.g. accumulation of toxic metabolites or altered physical properties of structural proteins)
   - the frequency of occurrence for this disease and whether it is specific to certain populations
   - what type of mutations are most commonly associated with the disease (e.g. frame-shift vs point)
   - is the mutation causing the disease a recessive or dominant trait (humans are diploids)
   - some of the gene’s physical characteristics—size, chromosomal location, is it part of an operon
   - is the gene product evolutionarily conserved in other organisms
   - is there and animal model to study the disease

Audience of the paper

You are to write this paper for a general audience and should assume that they do not have a strong background in science. This will require you to discuss the points listed above in a narrative format that explains the more technical aspects of your gene and gene product in a way that educates your reader. You can decide which technical aspects need explanation by asking whether or not you understand them yourself. If you don’t understand it, you can
bet your reader won’t either. Do not treat these points as a list of questions. Also, be careful to avoid “stitching” information from your references; this is easy to spot and is a form of plagiarism.

To place yourself in a position to facilitate writing the report, you might consider writing a pamphlet for families affected by the disease. If you chose this approach, you must still include the specific points listed above and layout the biochemical basis causing the disease, but you will need to do so in a fashion that is accessible to the non-scientifically trained. Don’t be afraid to include a diagram if you think it will help their understanding; think about how many times confusing written or verbal explanations have been made clear by a when accompanied by a diagram.

**Resources**

The primary resource you will use is the UCSC Genome Data Center. This site is only one of many for accessing genomic sequence information. One that I have found particularly good for getting to research articles is the NCBI PubMed database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=PubMed). If you decide to use this or any other database, it will take you some time to learn where in the database you will find the information you seek. With the explosion of information, databases have become structurally complex. The directions below will help you find the information you need in the USCS database.

**A note about naming conventions.** The abbreviations used for identifying genes and gene products (i.e. proteins) have a specific format or syntax. Both are usually given a three letter abbreviation, similar to your initials, which is followed by a letter or number identifier. The name of a gene is not capitalized but is italicized. The encoded proteins is capitalized but is not italicized. For example, when I was a grad student, I sequenced a gene encoding the pyolutein halogenase (*pltE*). The gene product, PltE, represented a new class of enzymes that can add a chlorine atom to an organic compound. Remember that as for any abbreviation, the first time you refer to a gene or gene product, be sure to use the complete name and define its abbreviation.

**USCS Database Introduction**

2. Find the link to the Blat database search tool. There are three access links.
3. On the new screen, make sure you are searching the Human genome; then type or cut/paste your sequence into the box and click the “submit” button.

![BLAT Search Genome](image)

4. A new screen will appear. Click on the “Browser” link.

![Human BLAT Results](image)

5. The window should show you a graphic representation of where in the human genome your sequence is located in addition to other lines representing the fact you have found database sequences that match. Click on the first line (labeled with 3 or 4 capital letters) representing a homologous sequence.
6. There are two menus in this view, a “Page Index” (links to places on the same page) and a “Sequence and Links to Tools and Databases” (links to external sources). These links will provide you with an endless supply of information related to your original sequence, including protein structure, homologous sequences in other organisms, and expression data. Take some time to discover what some of this data looks like.

For this assignment, you are most interested in protein information. The “UniProt Comments” link will take you to a location further down the page that provides an overview of information associated with this protein. If you are fortunate, there will also be a hyperlink to the Online Mendelian Inheritance in Man (OMIM) database in PubMed. The OMIM may also exist in other locations and have the format [MIM:####].

7. That should send you well on your way to completing this assignment. If not, please see me and we’ll get you set up.

**Searchable Sequences**

<table>
<thead>
<tr>
<th></th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GTGACGGAGAAAAATGGAGAAGGGTACGGGACGCGCCCTGTCAGATTTCACACCCGGATCCCCTGGCTCATGTCGGCAGGTGTCAGATCGTTCAGGCTGTCTGCTAGATCTGTGGTTTTTCTACTTGAAGGACACAATGTTTTCCAAACTAGCACATTTGCAGAGGTTTGCTGTACTTAGTCGCGGAGTTCATTCTTCAGTGGCTTCTGCTACATCTGTTGCAAACTAAAAAAACAGTCCAAGGCCCTCCAACCTCTGATGACATTCTTTGAAAGGGAATATAAGTATGGTGCACTGCAGGGCCAGGACGCACGTTTCAAGCTCATCAGTAAAGGTTCCTTAAATTCCCGAAGCAAGAAGTTAACAAGTAAAACAGCATCGGAACACCAGGATCCCATGACAGATTCTGTTGTCACGTCTCCTTACAGAGTTTGAGCGGTGCTGAACTGTCAGCACCATCTGTCCGGTCCCAGCATGCCTTCTGAGACCCCCCAGGCAGAAGTGAGTGATTACCCCGGGAGACTGAACTGCAGGGTTCAAGCCTTCCAGGGATTTGCCTCACCTTGGGGCCTT</td>
</tr>
</tbody>
</table>